Creating a Basic ACCESS Database

by Karen Stechuchak, MS and
Amy Jeffreys, MStat

ICER Biostatistics Unit

Http://hsrd.durham.med.va.gov/Biostat/


Outline of Presentation

- Motivation
- Components of a basic ACCESS Database
- Creating Data Tables
- Creating Data Entry Forms
- Enhancements
- Access ⟹ SAS
- Conclusion

Motivation

Good database design helps with….

- Consistent formatting for data entry
- Minimizing time spent on data cleaning
- Transferring to analysis applications with greater ease
- Formatting of data to be more amenable for analyses
- Documentation

Database Components

Data Tables
- Variable Name
- Data Type
- Descriptions and Captions
- Validation Rules
- Field Size
- Formats

Data Entry Forms
- Formatting
- Tab order
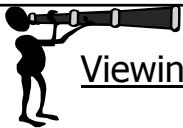- Field List

Other Components
- Queries
- Reports
- Macros

Viewing an ACCESS Table

In the design view of a table you can…
- designate variable names
- assign variable characteristics
- program simple data quality checks

In the datasheet view of a table you can…
- enter your data
- sort your data
- filter your data to only display records that meet a certain criteria

Variable Names

- Make SAS 6.12 compatible
  - <= 8 characters
  - no spaces (use underscore instead)
  - begin with a character
  - if longitudinal data, use <= 7 characters
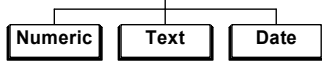- ICER Biostat Unit has current listing of standardized variable names

Data Types

3 Main Data Types

Numeric	Text	Date

Data Types to Use with Caution
- Yes/No
  - Can't distinguish between 'No' and missing data.
  - Internally recorded as 0 and -1 but when exported to other applications show True/False, T/F, 0/1 formats
- Memo Fields
- Time ⟹ Do not export well into SAS 6.12

Numeric Data Type

- Select data type as numeric whenever possible - much easier to analyze.
- Can format numbers for particular responses with lookup box
- If decimal values are allowed, choose field size to be single or double
- If whole numbers, choose integer
- Eliminate default value of 0

Lookup Boxes

- Appropriate for numeric data types with finite number of possible responses.
- Helps ensure consistent data entry
- Assists in mapping numeric codes to their description
- Use the lookup wizard under data type
- Different ways for lookup boxes to get its values

Creating Lookup Boxes

- Lookup Boxes from table/query
  - Use when have many variables with the same possible responses or when need to continuously add new response choices
- Lookup Boxes "typing in values"
  - Use when many variables have different response patterns and when response values are set (especially for survey data)
  - ICER Data Entry system currently allows SAS format program to be generated
- First visible column is how data is entered and displayed, but not necessarily stored

Text Data Type

- Avoid when possible - use numeric when finite number of responses
- Hard to analyze
- Time consuming to data enter
- Field size can range from 1-255 (default is 50)

Date Data Types

- Select date/time as data type
- Most common format is short date.
- Can use input mask to record 2 or 4 digit year
  - 4 digit year - input mask is 99/99/0000;0
  - 2 digit year - input mask is 99/99/00;0
- 4 digit year is safer - 1930 cutoff for 2 digit years
  - If entered 9/17/17, ACCESS would assume 9/17/2017, not 9/17/1917

Input Masks

- Ensures consistent and easier data entry
- Input masks only exist for text and date data types.
- For dates, prefer 99/99/0000;0
- Input masks for SSN, phone number, and zip code exist.
- Use caution with input masks for time (may not export well to some applications). If problems, record time in 2 numeric fields (one for hour and one for minute).

Descriptions and Captions

- Always enter a description of the variable
  - ICER Data Entry System creates SAS labels program using the description field
- Captions are optional
  - Can copy description as the caption
  - Will show up as column header in the table 'datasheet view' and also as label on the form.
  - If no caption is entered, the variable name is the column header in the table and label on the form.

Validation Rules

- Ensures data accuracy by prohibiting entry of out-of-range values
- Catches key stroke errors
- Prior to entering data…
  - decide on possible range of values for each variable
  - determine whether a null response is appropriate for any variables
  - program validation rules into design view of table

Validation Rules, Cont.

- Only include 'or Is Null' if possibility that data is not obtainable or is part of a skip pattern, or data entry within a record may be staggered.
- Use missing data response codes (i.e. 777 for don't know, 888 for refused, and 999 for missing data)
- Can use validation rules on numeric, character, and date data types

Validation Rule Examples

- If finite number of numeric choices*: 1 or 2 or 3 or 4 or 777 or 888 or 999
- Weight of adult patient: (Between 75 and 500) or Is Null
- Gender: 'M' or 'F' or Is Null
- Date: Between #01/01/1997# and #12/31/2000#

* If using A2K, you must change the Required property to 'Yes' to force entry of a response.

Primary Key

- Always make the unique ID number a primary key.
- Ensures that multiple people don't get the same ID number
- If you have longitudinal data, primary key a second variable.
- Example: entering procedures for a patient, create a variable that counts the procedure number being entered (1, 2, 3,…)
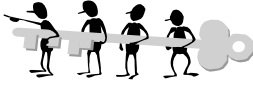
Creating a Primary Key

- Highlight the row(s) of the variables that you want to primary key.
- Then select Edit ⟹ Primary Key
- For longitudinal data, change your second indexed field to 'Yes (Duplicates OK)'.

Data Entry Form

- Makes data entry easier
  – helps you keep your place
  – less strain on eyes
  – can customize
- Do not create a form until your data table is completely finalized

Creating a Data Entry Form

- Click on the Forms tab and select New.
- Select Form Wizard and the table that the data will be stored in.
- Select the variables you want on the form (usually all of them).
- Choose a layout and style that you like best
- Name the form the same as the corresponding table
- Select Finish.

Customizing the Form

- Select Design View to position the controls as you wish.
- Each control has a label and input box, to move both click on the input box and drag.
- Try to structure the form as close to the hardcopy form as possible
- To change control properties, click on the control and select View ⟶ Properties
- Switch to Form View to test your form.

Tab Order

- Order that the cursor will move from control to control (variable to variable).
- After controls are positioned, check the tab order
- Select View ⟶ Tab Order.
- To change the order, select the row of variable to move and drag the row to appropriate place.
- Test by entering "dummy" data into the form

Field List

- List of the variables in the corresponding table
- If need to add or rename a variable in the table after the form has been created…
  - Select View ⟶ Field List
  - Highlight the variable.
  - Drag onto the form
  - Correct the tab order

Enhancements

- Queries
  - use to view, change, & analyze data
- Reports
  - present data in a printed format
  - customize appearance
- Macros
  - set of actions that perform particular operations
  - can automate tasks
- ICER Data Entry System

Access to SAS

- Access 97 tables to SAS 6.12
  - Export each data table as dbase V file
  - In SAS, use Import Wizard selecting the file format dbase table
- Access 97 tables to SAS 8.0
  - In SAS, use Import Wizard selecting the data source Microsoft Access 97 table

Conclusion

- Good table design on the forefront helps ensure better data quality in the end
- Select variable names and data types that are easier to analyze. (i.e. SAS 6.12 compatible)
- For more information on the ICER Data Entry system and list of standardized variable names, see ICER Biostatistics Unit